

Research Article

Research on the Mental Health of College Students Based on Fuzzy Clustering Algorithm

Qinghua Tang,¹ Yixuan Zhao ,² Yujia Wei,¹ and Lu Jiang¹

¹School of Public Health and Management, Guangxi University of Chinese Medicine, Nanning 530200, Guangxi, China

²Far Eastern University of the Philippines, Manila, Philippines

Correspondence should be addressed to Yixuan Zhao; 2019163761@feu.edu.ph

Received 29 July 2021; Revised 22 August 2021; Accepted 26 August 2021; Published 3 September 2021

Academic Editor: Jian Su

Copyright © 2021 Qinghua Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mental health of young college students has always been a social concern. Strengthening the supervision of college students' mental health problems is an important research content. In this regard, this paper proposes to apply fuzzy cluster analysis to the health analysis of college students and explore college students through fuzzy clustering. Explore the potential relationship between the factors that affect the health of college students, and this will provide a reference for the early prevention and intervention of college students' mental health problems. In view of this, an improved fuzzy clustering method based on the firefly algorithm is proposed. First, the Chebyshev diagram is introduced into the firefly algorithm to initialize the population distribution. Then, an adaptive step size method is proposed to balance exploration and development capabilities. Finally, in the local search process, a Gaussian perturbation strategy is added to the optimal individual in each iteration to make it jump out of the local optimal. The process has good optimization capabilities and is easy to obtain the global optimal value. It can be used as the initial center of the fuzzy C-means clustering algorithm for clustering, which can effectively enhance the robustness of the algorithm and improve the global optimization ability. In order to evaluate the effectiveness of the algorithm, comparative experiments were carried out on four datasets, and the experimental results show that the algorithm is better than the comparison algorithm in clustering accuracy and robustness.

1. Introduction

Clustering technology divides datasets into different categories and data between categories. The differences are large, and the differences within the same category are as small as possible. According to the National Center for Disease Control and Prevention and the Center for Mental Health, survey data show 10%–25.4% of college students are interested in questions. The former State Education Commission has conducted psychological tests and investigations on 12,600 college students, and 20.23% of them had obvious psychological problems. Every year in the university, more than 10 students take their own humiliation in China due to broken relationships, interpersonal tensions, frustrations in life, etc. It is estimated that psychological problems are leading to college students taking a break from school, dropping out of school, and doing other evil things. It can be

seen that psychological problems have a serious influence on college students' health growth [1]. To strengthen the psychological health education of college students, and to cultivate the moral and physical health of students, is the new task of high school education. In recent years, the psychological health of college students has become a focal point of the whole society due to the frequent occurrence of evil incidents caused by the psychological problems of college students. Experts at home and abroad have also done a lot of in-depth research on this and actively explore college students' psychological guidance and intervention model. Some developed countries in the west attach great importance to the mental health education of young students [2]. They analyze the psychological data of students by means of measurement, mathematical statistics, methods, and tools. In order to guide psychological teaching assistants to give students psychological guidance with acupuncture, they

need to conduct rigorous training [3]. In the past, the emphasis on the psychological health education of Chinese university students has been increasing. At the same time, the education departments at all levels have done a lot of work around the psychological health education of college students. At the same time, through a variety of assessment measures, understand the importance of mental health education in the management of high school students and quality education [4].

However, there are still many problems in the development of mental health education in domestic colleges and universities, which need to be further strengthened and improved, such as the uneven development of various colleges and universities. The relevant institutions for the test are not standardized enough, the psychological test data are relatively single, and the tutor resources are not sufficient enough. In this paper, the model clustering analysis should be used in the analysis of college students' mental health, through the model clustering, to analyze the rules and characteristics of college students' mental health and the behavior characteristics of students with mental disabilities [5]. Explore various potential relationships between psychological test data, and provide scientific basis for universities to establish and improve early prevention and intervention of students' psychological barriers [6].

Cluster analysis is an important technology in the field of data mining. By using it to find the similarity and difference between data, we can further mine the potential relationship between data [7]. With the deepening of understanding and the needs of practical application, the boundaries between real data objects are often not obvious, not an either or relationship [8]. They have a certain fuzziness. If the clustering analysis method is used to realize the classification of some data, it will cause great errors [9].

Fuzzy clustering analysis provides the possibility for the classification of fuzzy things. In 1966, Bellman, Kalaba, and Zadehw first proposed to deal with clustering problems based on fuzzy sets. Later, Wee, Flak and Turner, Gitman, and Levine also used fuzzy sets to study classification problems [10]. In 1969, Ruspini defined the concept of fuzzy partition of dataset for the first time; fuzzy clustering is systematically expressed and studied [11].

Scholars try to study fuzzy clustering from many aspects and also put forward many fuzzy clustering methods, such as the relaxed k-nearest neighbor method proposed by Taogu and Bubulsson. Leahy put forward the clustering method of optimal graph theory, Kler et al. [12] proposed a fuzzy k-nearest neighbor method, and Bezdek et al. proposed a fuzzy k-nearest neighbor method, esogbue's dynamic fuzzy clustering method [13].

The most widely used method in practical application is the fuzzy clustering method based on objective function first proposed by Ruspini. Later, Duim and Bezdek extended the hard C-means clustering algorithm to the fuzzy state [14], established the fuzzy C-means clustering theory, verified the convergence of fuzzy C-means (FCM) algorithm, and really proposed an effective method. So far, scholars have deeply explored the fuzzy clustering algorithm based on objective function from various aspects, such as improving the

convergence speed of FCM algorithm, modifying the objective function and fuzzy partition space, and fuzzy clustering of noisy data and mixed data. The convergence and stopping criteria of the FCM algorithm are one of the important characteristics, the fuzzy clustering algorithm is sensitive to initialization and difficult to achieve global optimization, and remarkable results have been achieved [15].

The amount of data stored in the University Yeli health database is increasing. These data contain a lot of information that can prevent students' Yeli barriers in the early stage. Timely, efficient, accurate, and systematic analysis of this information will greatly improve the efficiency of college students' health education [16]. Fuzzy cluster analysis has a good application in three aspects. It will integrate relevant technologies and provide a scientific reference for practical decision-making.

The main contributions of this paper are as follows:

Based on the classical fuzzy C-means (FCM) clustering algorithm, a FCM algorithm based on information entropy attribute weighting is proposed, which solves the defects of FCM algorithm, such as very sensitive to initialization, easy to fall into local convergence, unable to achieve global optimization, and unable to deal with under-regular datasets. Firstly, the information entropy theory is introduced to initialize the cluster center [17]. Secondly, the idea of class merging is used to continuously adjust the position of the initial cluster center and determine the final clustering center. Finally, the idea of attribute weighting is introduced to solve the problem of local convergence and achieve the global minimum. The improved algorithm will greatly improve the efficiency and accuracy of the original algorithm [18].

The research innovations of this paper are as follows:

- (1) This article proposes to apply fuzzy clustering analysis to the health analysis of college students and to explore the laws and characteristics of college students' health through fuzzy clustering.
- (2) An improved fuzzy clustering method based on the firefly algorithm is proposed.
- (3) The Chebyshev diagram is introduced in the firefly algorithm to initialize the population distribution. Then, an adaptive step size method is proposed to balance exploration and development capabilities.

The fuzzy clustering algorithm based on information attribute weighting is applied to the analysis of college students' mental health data. Firstly, the theoretical rationality of the algorithm is analyzed, and the specific data analysis test is performed. The advantages of the algorithm compared with the fuzzy C-means algorithm are analyzed through experiments. At the same time, the class structure hidden in the mental health data attribute is excavated, and the main factors affecting college students' mental health are analyzed, which is convenient for the school to analyze and adjust the work ideas of private mental health education and formulate corresponding psychological intervention measures [19].

2. Analysis of College Students' Mental Health

2.1. Mental Health of College Students. Freshmen studying in Xinyang University are in a critical period of shaping and forming life values, and it is also a period of rapid growth and development. At this time, different steps of psychological and physiological growth and social change and development are bound to induce psychological obstacles. Understanding the overall mental health status of freshmen in colleges and universities is of great significance to comprehensively promote students' quality education.

2.2. Mental Health Implications. Due to the various forms of mental health, there is no unified standard for mental health. Based on various studies, psychologists believe that mental health includes the following seven aspects: normal intelligence, emotional health, sound will, behavior coordination, interpersonal adaptation, moderate response, and psychological characteristics in line with age.

2.3. College Student Personality Questionnaire. University Personality Inventory (UPI) is a kind of mental health questionnaire for college students, which aims to find and treat students with psychological problems early. The personality questionnaire of college students is mainly used as a specific survey of the mental health status of freshmen when they enter school, so as to better understand the psychological and mental health status of students and help psychological counseling teachers find students with psychological problems as soon as possible and timely provide targeted psychological guidance and corresponding psychological crisis intervention measures for these college students.

The classification and processing of UPI results are combined with test results and actual diagnosis; students are divided into three categories: A, B, and C, which respectively represent obvious psychological abnormalities, general psychological problems, and normal psychology. How to judge the three categories A, B, and C is mainly based on the experience of psychological counselors.

2.4. Symptom Self-Assessment Form. SCL-90 symptom checklist is the most widely used scale in mental health testing and measurement. The scale was compiled by Derogatis of Germany in 1975. It contains 90 questions. The contents of the questions range from simple and basic to trait changes. It comprehensively evaluates a person's psychological symptoms and their severity. It has the advantages of large amount of data, wide coverage of symptoms, and more accurate and comprehensive description of the subjects' conscious symptoms.

The scale includes 90 items and 9 subscales in total. Each item of the scale is scored at level 1–5 (no, very light, medium, heavy, and serious). There is no rigid provision on the specific meaning.

A total of 4186 students of grade 2015 of Xinyang University participated in the psychological test, including

students from eight departments: Foreign Language Department, Chinese Department, Mathematics and Computer Department, Science and Engineering Department, Economic Management Department, Social Science Department, Wood Engineering Department, and Art Department. 4161 valid test data are mainly referred to this time.

2.5. Data Preprocessing

2.5.1. Data Selection. In the mental health data obtained from the test, due to the uniqueness of attribute values such as student number, gender, and test answer, the grade is 2015, and the test is a unified date; these attributes have no impact on the clustering results and only increase the complexity of calculation. Delete some attributes directly during clustering.

2.5.2. Data Cleaning. Data cleaning is to deal with the missing value of the data part. For example, in the case of a small amount of data, add a value to the missing value according to the distribution law of the data value. In the case of a large amount of data, W can be replaced by a constant value. For some students' basic information, such as family name and department, the missing value of the genus can be substituted by 0. After the data cleaning, delete the invalid data and replace the useful missing data.

2.5.3. Establishing the Database and Data Table for College Students. In the process of cluster analysis, data collection is the first step, followed by data preprocessing and clearing up invalid data. Establish a database of psychological well-being for college students. The data of this paper are obtained from the personality test of 2015 college students of Xinyang University and the Symptom Self-Rating Scale of CSX-90, in which the symptom self-rating scale set contains ten factors. Each factor reflects a particular aspect of a person's symptoms.

2.5.4. Algorithm Implementation Process. College students' health data analysis: the first is to collect college students' health data, through the selection and cleaning of data attributes and integrate effective data to establish a comprehensive private health testing database. The data were clustered using the FCM algorithm based on the attribution weighting of information. The clustering analysis process of college students' health data is shown in Figure 1.

3. IFAFCM Algorithm

3.1. FCM Algorithm. FCM is a common clustering algorithm that distributes sample data according to their membership degree. Its objective function formula is shown in the following formula:

$$J_{\text{FCM}}(U, V) = \sum_{i=1}^c \sum_{j=1}^n u_{i,j}^m d_{ij}^2. \quad (1)$$

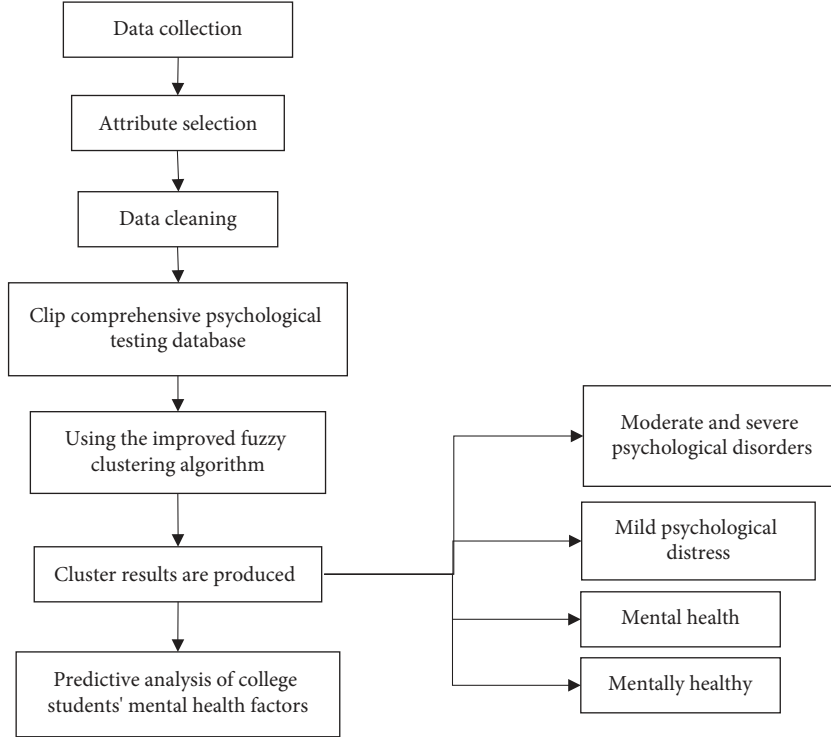


FIGURE 1: The clustering analysis process of college students' health data.

Here, n represents the total number of samples, and its sample dataset $X = \{x_1, x_2, \dots, x_n\}$, $x_j = \{x_{j1}, x_{j2}, \dots, x_{js}\}$ represents s attributes of the j -th sample; C represents the number of cluster centroid, $V = \{v_1, v_2, \dots, v_c\}$ represents the centroid set; element u_{ij} represents the size of the j -th sample belonging to the center of mass I , and m represents the fuzzy index, usually taking a constant value of 2; and d_{ij} is the Euclidean distance from each sample to the cluster center. The membership matrix u_{ij} and the centroid matrix v_i are calculated as shown in the following equations:

$$u_{ij} = \frac{1}{\sum_{k=1}^C (d_{ij}/d_{kj})^{2/m-1}}, \quad (2)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}. \quad (3)$$

3.2. FA Algorithm. As a swarm intelligence optimization algorithm, FA has more advantages than particle swarm optimization and genetic algorithm in global optimization [17–19]. The algorithm searches and optimizes fixed space by simulating the process of attraction and movement between fireflies in nature. The firefly makes the darker individuals move to the brighter ones through its own fluorescein luminescence, and finally all the individuals tend to the brightest firefly, and its position is the optimal position in the search space. The luminance, attractiveness, and position update formulas for fireflies are described by the following definitions.

Definition 1. The relative fluorescence of two fireflies is shown in the following formula:

$$I = I_0 e^{-\gamma r_{ij}}. \quad (4)$$

Here, I_0 represents the highest fluorescence brightness obtained when $\gamma = 0$, and the better the objective function value is, the higher the brightness will be; γ represents the absorption coefficient of light intensity as the distance between the two fireflies increases, indicating that the fluorescence is gradually weakened by the air medium; and r_{ij} represents the distance between firefly i and firefly j . The calculation formula is shown as formula (5), and s represents the dimension:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{l=1}^s (x_{il} - x_{jl})^2}. \quad (5)$$

Definition 2. The attraction degree of fireflies is shown in the following formula:

$$\beta = \beta_0 e^{-\gamma r_{ij}^2}. \quad (6)$$

Here, β_0 represents the maximum attraction, that is, the attraction when $r = 0$, γ is the same as above.

Definition 3. The firefly position update formula is shown as the following formula:

$$s_i(t+1) = s_i(t) + \beta(s_j(t) - s_i(t)) + \alpha(\text{rand} - 0.5). \quad (7)$$

Here, s_i and s_j are the positions of firefly i and j in the search space, respectively; t represents the number of iterations; β is the attractiveness; α is the step factor, its value range is in the interval $[0, 1]$, where the value is a constant; and rand is the random factor, which is the random number on the interval $[0, 1]$.

3.3. Improved Firefly Algorithm

3.3.1. Chebyshev Maps. In order to verify the effectiveness of Chebyshev chaotic mapping, the Chebyshev chaotic mapping method was used to evenly distribute 20 two-dimensional particles into the three-dimensional plane formed by function z (as shown in Figure 2) and compared with the three-dimensional graph of traditional randomly distributed particle swarm optimization (as shown in Figure 3). The expression of function z is shown in the following formula:

$$z = 3(1-x)^2 e^{(-y^2-(x+1)^2)}. \quad (8)$$

By comparing Figure 2 with Figure 3, it can be seen that Chebyshev mapping method can evenly distribute the initial firefly population in the search space. Under the same conditions, particles generated by Chebyshev mapping can collect more information, effectively avoiding the problem of particle swarm falling into local optimal. Compared with the commonly used chaotic mapping logistic mapping, Chebyshev mapping has more advantages in distribution range and distribution mode.

In this paper, the Chebyshev mapping method is adopted to generate chaotic variables, and the formula is shown as follows:

$$x_{n+1} = \cos(k \arccos x_n), x_n \in [-1, 1]. \quad (9)$$

Here, k is the order and can generate an unrelated chaotic sequence with ergodic property. The steps for Chebyshev to map the population number of fireflies are as follows:

- (A) Define the population number N of firefly and randomly generate a D -dimensional vector as the first firefly individual in the D -dimensional space, $Y = (y_1, y_2, \dots, y_D) \forall i \in [-1, 1], i \in [1, d]$;
- (B) $N-1$ iterations were performed for each dimension of Y using formula (9) to generate the remaining $N-1$ firefly individuals;
- (C) The mapping formula is shown in formula (10), which maps the generated N firefly individuals to the solution search space in turn:

$$x_{id} = \frac{l_d + (1 + y_{id}) \times (u_d - l_d)}{2}. \quad (10)$$

Here, u_d and l_d are the upper and lower bounds of the d -th dimension of the search space, respectively; y_{id} is the d -th dimension of the i -th firefly; and x_{id} is the coordinate value of the d -th dimension of the i -th firefly in the search space.

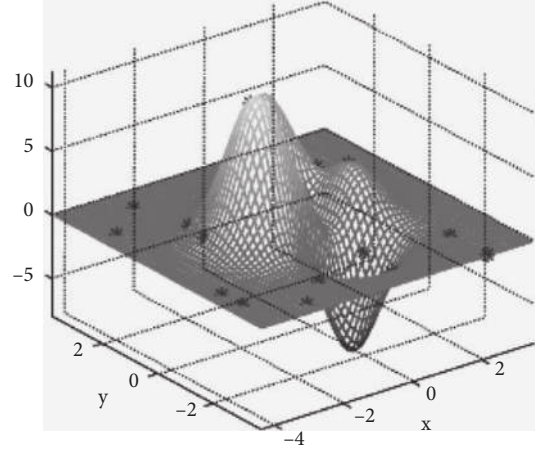


FIGURE 2: Chebyshev chaotic map distribution.

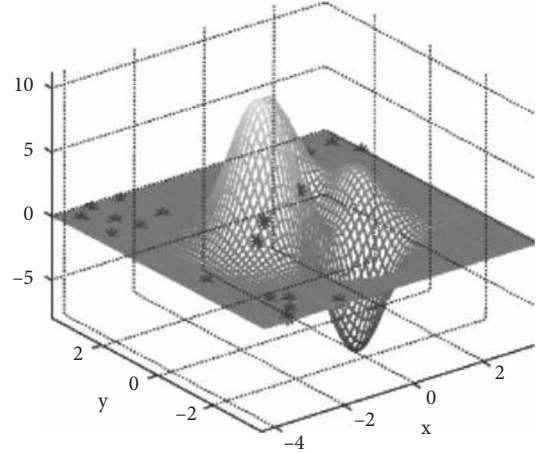


FIGURE 3: Traditional random distribution.

3.3.2. Adaptive Step Factor. The optimization ability of firefly population is mainly reflected in the individual position updating method. In formula (7), the first term represents the current location of fireflies; the second item indicates the global search ability of fireflies; the third item represents the local search ability of fireflies, in which the adjustment of step factor is very critical.

Parameter α in FA algorithm balances the ability of firefly population exploration and development, and the value range of α is $[0, 1]$. Too large or too small random value may seriously affect the search ability of the population. Therefore, an adaptive step mechanism is designed to enable the population to independently and timely change the global search and local search in the search process, so as to improve the ability to find the optimal solution. The calculation formula of the control parameter α is shown as the following formula:

$$\alpha_{t+1} = \alpha_t \delta, \quad 0.5 < \delta < 1. \quad (11)$$

Here, δ represents the cooling coefficient. At the early stage of iteration, a larger value is given to the step factor α to ensure a better global optimization capability in the early

stage. After that, the α value decreases gradually in the iterative process, ensuring better local search capability. By adjusting the α value, the global search ability and local search ability of firefly algorithm can be better balanced.

3.3.3. Gaussian Disturbance Mechanism. When FA algorithm falls into the local optimal state, its own random disturbance strategy cannot jump out of the local optimal position under multidimensional conditions, so a more adaptable disturbance mechanism is needed. Gaussian perturbation is a commonly used perturbation strategy for local optimal problems. It carries out Gaussian perturbation on the optimal individual obtained after each iteration to obtain a new optimal individual until the Gaussian perturbation cannot be obtained.

The new optimal individual indicates that the algorithm has reached the global optimal.

The basic principle of Gaussian perturbation is to add a small amplitude random vector to the position vector of the optimal individual in each iteration process, and this vector should conform to the normal distribution. The formula is as follows:

$$Y_i^{t+1} = Y_i^t \times [1 + N(0, 1)]. \quad (12)$$

Here, Y_i^t represents the current position of the i -th firefly; Y_i^{t+1} represents the new position formed by Gaussian disturbance; and $N(0, 1)$ is a random vector and obeys normal distribution.

In each iteration of the algorithm, Gbest corresponds to the contemporary optimal position and is subjected to Gaussian perturbation. The formula is as follows:

$$NGbest = Gbest(1 + N(0, 1)). \quad (13)$$

Vector NGbest represents the new position vector formed after Gaussian disturbance, and then the global optimal position is updated. The update formula is shown as the following formula:

$$Gbest^{t+1} = \begin{cases} NGbest^t, & \text{if } (NGbest^t < Gbest^t) \\ Gbest^t, & \text{otherwise} \end{cases}. \quad (14)$$

To sum up, the steps of the improved firefly algorithm are shown in Figure 4.

4. Experimental Results and Analysis

4.1. Experimental Operation Platform. The program running environment of this experiment is processor CPU Intel Core I7-8750H, main frequency 2.20 GHz six-core, memory 8 GB, operating system Windows 10 64bit, and integrated development environment for MATLAB 2014A.

4.2. Experimental Dataset. During the experiment, four datasets commonly used in UCI database 1 were adopted: Iris dataset, Ecoli dataset, Wine dataset, and Glass dataset; the detailed information of each dataset is shown in Table 1. In addition, in the process of clustering, in order to avoid the

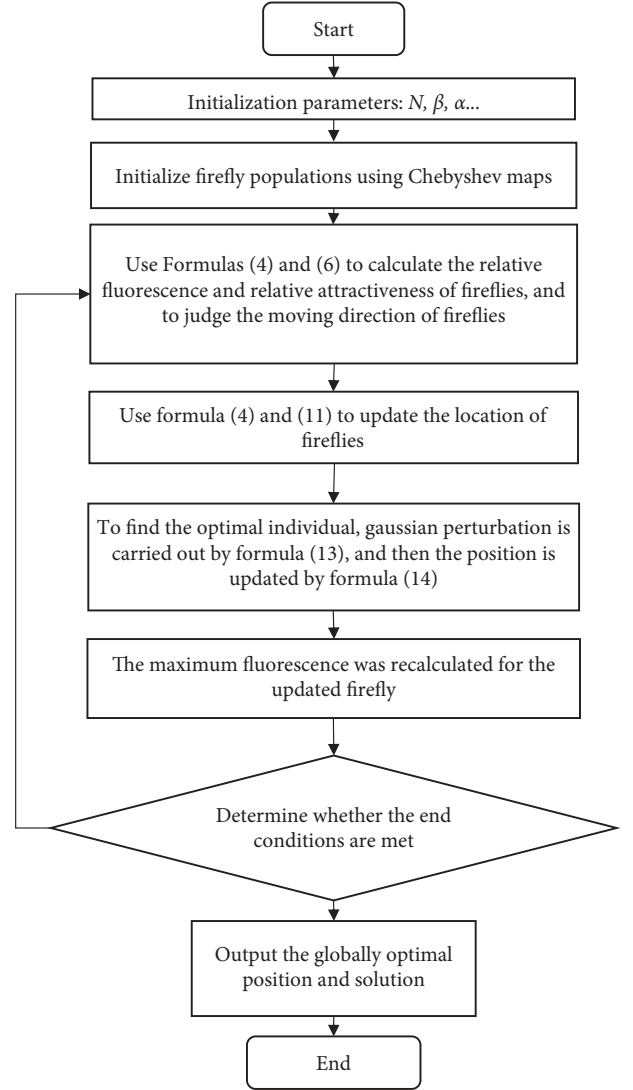


FIGURE 4: Flow chart of improved firefly algorithm.

influence of the clustering results caused by the great difference between the characteristic values, the dataset values are normalized so that the characteristic values are between $[0, 1]$.

4.3. Evaluation Indicators. In this experiment, clustering accuracy (ACC) and normalized mutual information (NMI) were used to evaluate clustering performance. Clustering accuracy: the ratio of the number of samples correctly assigned to clustering to the total number of samples requiring clustering. The calculation formula is as follows:

$$ACC = \frac{\sum_{k=1}^K d_k}{N}. \quad (15)$$

Here, DK represents the number of samples correctly assigned to the K th cluster, and N represents the number of all samples.

Standardized mutual information: a symmetric metric that quantifies statistical information shared between two cluster distributions. The measurement formula is as follows:

TABLE 1: Datasets used in the experiment (number of people).

Datasets	Number of samples	Number of attributes	Number of categories
Iris	150	4	3
Ecoli	336	7	8
Wine	178	13	3
Glass	214	10	7

TABLE 2: Comparison of the results of six algorithms under three test functions.

Function		FA [14]	FAGO [14]	FABE [14]	PSO [19]	GA [19]	IFA
F1	AVG	2.00E-03	1.74E-04	4.51E-04	1.83E+04	1.03E+03	1.21E-08
	STD	7.03E-05	1.35E-05	9.24E-06	3.01E+03	5.79E+02	5.43E-08
F2	AVG	1.198E+02	1.97E+01	5.28E+00	1.96E+07	1.94E+04	2.59E-01
	STD	3.83E+01	2.90E+01	1.62E+01	6.25E+06	1.31E+04	3.97E-01
F3	AVG	2.60E-01	2.37E-04	1.07E-05	1.70E+02	1.01E+01	1.67E-11
	STD	3.57E-04	6.23E-09	1.36E-08	3.17E+01	2.43E+00	7.43E-11

$$\text{NMI}(R, Q) = \frac{\sum_{i=1}^I \sum_{j=1}^J P(i, j) \log P(i, j) / P(i)P(j)}{\sqrt{H(R)H(Q)}}. \quad (16)$$

Here, R and Q are the two partitions of the input dataset, including cluster i and cluster j ; $P(I)$, respectively, is the probability of being assigned to cluster R_i from randomly selected samples in the dataset; $P(i, j)$ is the probability of the sample belonging to cluster R_i and cluster Q_j at the same time; and $H(R)$ is the entropy associated with all the probabilities $P(I)$ ($1 \leq i \leq I$) in partition R . The higher the accuracy value in the experimental results, the more the sample is correctly allocated to the corresponding cluster, and the better the clustering performance. The larger the value of standardized mutual information is, the more accurate the partition between clusters is, and the better the clustering effect of the algorithm is.

4.4. Experimental Result. The proposed algorithm was used to run 30 particles in three test functions for 20 times, each round of iteration was 500 times, and its average value and standard deviation were calculated and then compared with the other five algorithms: FA [14], FAGO [14], FABE [14], PSO [19], and GA [19]. The comparison results are shown in Table 2. It is proved that the algorithm has good competitiveness in finding the most value point among the six kinds of algorithm precision single-peak function or multi-peak function, and it can solve the problem of FCM algorithm falling into the local center of the initial cluster center to a certain extent. It can provide a solid basis for the results of the improved fuzzy clustering algorithm. Comparison of the results of six algorithms under three test functions is shown in Table 2.

It can be seen from Table 2 that the simulation results of the cluster analysis algorithm have better simulation results than other models, and the accuracy is higher, and it has better trend analysis for the researched problems.

4.4.1. Comparison Experiment of Clustering Performance. The algorithms FCM, PSOFM, FAFM, and IFAFM were compared on four datasets. The parameters of the

TABLE 3: Comparison of the results of evaluation indexes ACC and NMI of the four algorithms, %.

Algorithm	Iris		Ecoli		Wine		Glass	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
FCM	89.33	74.04	52.38	48.29	94.38	82.99	38.79	30.41
PSOFM	90.67	75.92	54.36	49.65	94.71	83.24	42.19	33.66
FAFM	91.13	78.84	58.93	51.77	95.51	84.73	44.99	36.44
IFAFM	93.12	84.62	62.93	56.71	96.35	87.46	49.63	45.21

algorithm were set as the maximum iteration number $\max T = 300$, fuzzy index $m = 2$, threshold $\varepsilon = 10^{-5}$, initial population $N = 30$, maximum attraction $\beta_0 = 1.0$, step factor $\alpha = 0.8$, cooling coefficient $\delta = 0.98$, light absorption factor $\gamma = 0.9$, and particle acceleration coefficient $C1 = c2 = 2$; $r1$ and $R2$ are random numbers distributed in $[0, 1]$; inertia weight coefficient $W_{\max} = 0.9$, and $W_{\min} = 0.4$.

Each algorithm was run on each dataset for 10 times, and ACC and NMI were obtained by the average value. The comparison results are shown in Table 3.

5. Conclusion

This paper proposes an improved fuzzy clustering algorithm IFAFM to improve the initialization sensitivity and poor robustness of traditional FCM, thereby improving the clustering accuracy. First, the Chebyshev chaotic map is used to evenly distribute the firefly population in the search space to generate an initial population with a wide range, uniform distribution, and strong global search ability. Then, the step factor is transformed into a dynamic adjustment mode to balance the exploration and development process. Finally, Gaussian disturbance is used to mutate the optimal individual to enhance the ability of the algorithm to jump out of the local optimum. The combination of the firefly algorithm and the traditional FCM algorithm not only provides a good global search capability but also ensures a good local development capability. In order to evaluate the clustering effect of the IFAFM algorithm, experiments were conducted on four datasets and compared with the other three algorithms. The experimental results show that compared

with the other three algorithms, the IFAFCM algorithm has improved clustering performance and antinoise ability. In future research, as the factors affecting the mental health of college students are becoming more and more complex, it is necessary to develop more parameters for analysis, which is an important topic for future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy C-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.
- [2] H. Zhang, Q. Wang, W. Shi, and M. Hao, "A novel adaptive fuzzy local information SC\$ -means clustering algorithm for remotely sensed imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5057–5068, 2017.
- [3] N. Chetih, Z. Messali, A. Serir, and N. Ramou, "Robust fuzzy c-means clustering algorithm using non-parametric Bayesian estimation in wavelet transform domain for noisy MR brain image segmentation," *IET Image Processing*, vol. 12, no. 5, pp. 652–660, 2018.
- [4] B. Dogan and M. Korurek, "A new ECG beat clustering method based on kernelized fuzzy C-means and hybrid ant colony optimization for continuous domains," *Applied Soft Computing*, vol. 12, no. 11, pp. 3442–3451, 2012.
- [5] R. J. Kuo, T. C. Lin, F. E. Zulvia, and C. Y. Tsai, "A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis," *Applied Soft Computing*, vol. 67, no. 6, pp. 299–308, 2018.
- [6] A. Yelghi and C. Köse, "A modified firefly algorithm for global minimum optimization," *Applied Soft Computing*, vol. 62, no. 1, pp. 29–44, 2018.
- [7] B. Wang, D.-X. Li, J.-P. Jiang, and Y.-H. Liao, "A modified firefly algorithm based on light intensity difference," *Journal of Combinatorial Optimization*, vol. 31, no. 3, pp. 1045–1060, 2016.
- [8] J. Senthilnath, S. N. Omkar, and V. Mani, "Clustering using firefly algorithm: performance study," *Swarm and Evolutionary Computation*, vol. 1, no. 3, pp. 164–171, 2011.
- [9] X. S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *International Journal of Bio-Inspired Computation*, vol. 2, no. 2, pp. 78–84, 2010.
- [10] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, no. 8, pp. 849–872, 2019.
- [11] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, no. 3, pp. 583–617, 2003.
- [12] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.
- [13] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1328–1337, 2010.
- [14] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 573–584, 2013.
- [15] G. M. Villeirs, G. O. De Meerleer, K. L. Verstraete, and W. J. De Neve, "Magnetic resonance assessment of prostate localization variability in intensity-modulated radiotherapy for prostate cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 60, no. 5, pp. 1611–1621, 2004.
- [16] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain, "Simultaneous feature selection and clustering using mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1154–1166, 2004.
- [17] R. R. Gharieb, G. Gendy, and A. Abdelfattah, "C-means clustering fuzzified by two membership relative entropy functions approach incorporating local data information for noisy image segmentation," *Signal, Image and Video Processing*, vol. 11, no. 3, pp. 541–548, 2017.
- [18] M. Du, S. Ding, and H. Jia, "Study on density peaks clustering based on k-nearest neighbors and principal component analysis," *Knowledge-Based Systems*, vol. 99, pp. 135–145, 2016.
- [19] J. Xie, H. Gao, W. Xie, X. Liu, and P. W. Grant, "Robust clustering by detecting density peaks and assigning points based on fuzzy weighted K-nearest neighbors," *Information Sciences*, vol. 354, no. c, pp. 19–40, 2016.